

An interactive 3D medical visualization system based on a light field display

Marco Agus · Fabio Bettio · Andrea Giachetti · Enrico Gobbetti ·
José Antonio Iglesias Guitián · Fabio Marton · Jonas Nilsson · Giovanni Pintore

Abstract We present a prototype medical data visualization system exploiting a light field display and custom direct volume rendering techniques to enhance understanding of massive volumetric data, such as CT, MRI, and PET scans. The system can be integrated with standard medical image archives and extends the capabilities of current radiology workstations by supporting real-time rendering of volumes of potentially unlimited size on light field displays generating dynamic observer-independent light fields. The system allows multiple untracked naked-eye users in a sufficiently large interaction area to coherently perceive rendered volumes as real objects, with stereo and motion parallax cues. In this way, an effective collaborative analysis of volumetric data can be achieved. Evaluation tests demonstrate the usefulness of the generated depth cues and the improved performance in understanding complex spatial structures with respect to standard techniques.

1 Introduction

Most of the digital diagnostic datasets acquired by modern imaging modalities, such as CT, MRI or PET scanners, are intrinsically three-dimensional. Until very recently, these datasets have, however, been primarily analyzed in radiology departments by displaying their 2D cross-sections. Slice based techniques are well established and understood, and enable to rapidly implement operations such as dataset browsing. However, with these methods, physicians are forced to mentally combine multiple

views to construct a mental model of 3D shapes, adding further complexity to an already difficult task. Moreover, slice-based methods are reaching their usability limits, if only because of the sheer amount of slices produced by modern scanners. Both the 3D complexity and the enormous size of medical data thus create a growing demand for interactive high quality 3D visualization.

The rapid evolution of commodity hardware has recently made it possible to achieve high quality interactive volumetric rendering of diagnostic data on reasonably inexpensive platforms. Most modern radiology workstations now include 3D modules which can generate impressive virtual representations of the imaged structures. 3D visualization for medical analysis is not, however, widely accepted and applied as it could be. The problem is that the classical 3D radiology workstation, even if very effective, still suffers from relevant shortcomings limiting its impact. Typical medical datasets contain many overlapping structures, leading volumetric techniques to generate cluttered images, which may become difficult to understand when projected onto a 2D screen. Enhancing depth and shape perception in volumetric rendering is thus a very active research area, which is tackled from different angles. Recent contributions include methods for supporting real-time rendering and user interaction, improving rendering quality with advanced photorealistic models, or developing non-photorealistic approaches to emphasize model features by illustrative techniques. In general, however, the user needs a lot of interaction to obtain the correct point of view and/or to generate adequate motion parallax for improving depth perception. Metaphors used to navigate in 3D scenes with standard input devices are often not immediate for the radiologist. Furthermore, they increase cognitive load and often hinder good collaborative work, with just one person controlling

the navigation and others just looking at the screen from different positions.

An orthogonal research direction consists in improving volumetric understanding by presenting results on 3D displays able to elicit more depth cues than the conventional 2D monitor, and build around them environments that support collaborative discussion. The goal of such systems would ideally be to generate high quality spatial visualizations, enabling the user to perceive detailed virtual objects floating in space. Such 3D visualizations would enhance volumetric understanding, as there is evidence that ego and/or model motion as well as stereopsis are essential cues to achieve rapid direct perception of volumetric data [4, 16]. Moreover, in such environments, much of user interaction would be replaced by simple, natural head motions, reducing interface complexity and user burden. Most importantly, by providing consistent shared views of synthesized objects in the display workspace, multi-user discussion would be naturally supported.

In order to be effective in a radiological setting, such an approach should meet a number of additional requirements. First of all, it should not require users to wear complex tracking or display devices. Moreover, the visualization area should be large enough to allow several radiologists and physicians to examine the data in a collaborative way, much as they are used to do with 2D printed slides in front of an illuminator, avoiding the necessity of performing complex and unnatural navigation tasks. In terms of rendering quality, the system should be able to support various rendering techniques proposed for enhancing anatomy understanding, as well as for supporting arbitrarily large datasets, a fundamental issue when dealing with the increasingly growing resolution of digital diagnostic modalities.

In this paper, we present a prototype 3D medical data visualization system which meets these requirements, within certain limits, clearly discussed in this article. It is based on the combination of a light field display, i.e., a display supporting high resolution direction selective light emission, with customized advanced rendering techniques. Recent advances in 3D display design prove that high resolution display technology able to reproduce natural light fields is practically achievable using an advanced evolution of multiview display technology [2]. Rendering for such displays requires generating a large number of light beams of appropriate origin, direction, and color, which is a complex and computationally intensive task. Moreover, the optical characteristics of these displays impose specialized rendering methods. In previous work, we have shown that displays of this class can be used for complex data [3, 1] by developing customized hardware accelerated rendering techniques. In order to obtain all of the

other features required for a prototype 3D medical workstation, we integrated our methods into a DICOM environment, added a simple interaction system based on a 3D mouse, an optimized screen, virtual object and user positions in order to obtain the best perceived quality and the easiest multi-user interaction. Moreover, in order to support extremely large datasets, we customized a recently introduced [11] out-of-core accelerated direct volume rendering algorithm. The original algorithm has been further enhanced to fully support high resolution, 16 bits, raw medical datasets as well as labeled volumes. The renderer also supports several illustrative rendering methods. The advantages obtained with respect to standard 3D workstations have been assessed with psychovisual tests, which have proved the usefulness of the depth cues provided by the display and showed that it allows users to obtain better results in complex tasks involving spatial ability.

In the following sections, we discuss the implementation and evaluation of our prototype system. Sec. 2 gives a brief overview of light field display technology, Sec. 3 outlines the multiresolution direct volume rendering method developed for this class of devices and discusses the optimizations introduced for high quality rendering of high resolution medical data. Section 4 discusses the implementation and several issues related to the interaction design. Section 5 describes the evaluation tests performed with users and reports on a preliminary test of the diagnostic capabilities of the system. Sec. 6 discusses advantages and limitations of our prototype system and planned future work.

2 Light field display technology

2.1 3D display approaches

Immersive Virtual Reality environments, such as CAVE systems, have been tested in real diagnostic tasks, e.g., 3D echo analysis, and it was shown that they are capable of revealing defects that will escape 2D analysis and that they have a faster learning curve than traditional radiological workstations [18]. Current immersive VR systems can support only a very limited number of simultaneous users, most often a single one, and typically require head tracking and other hardware that does not fit well into the standard working environment of a radiology department. Several approaches have been proposed to support a real 3D visualization without the classical single point of view or single user limitations of traditional stereo displays, and many of them have been applied to medical data visualization. We provide here only a rapid survey of the subject, with a particular emphasis on the most closely related approaches. The key technical feature of 3D displays is direction-selective light emis-

sion, which is most commonly obtained by volumetric, holographic, or multiview approaches.

Volumetric displays synthesize light fields by projecting light beams onto refractive/reflective media positioned or moved in space (e.g., [15,9,20,13,6]). A display of this class, Actuality Systems Inc. Perspecta, has been used to visualize DICOM datasets and tested in a radiology planning system [10,17]. The main disadvantages are the limited scalability of the approach, and the difficulty in presenting occlusion effects. Because of mechanical constraints, these approaches are practical only for limited image sizes and model complexity. Pure holographic techniques are based on generating holographic patterns to reconstruct the light wavefront originating from the displayed object [21,22,12]. Although this approach can theoretically provide the most compelling imagery, generating dynamic images is very complex, and, in current prototypes, the hardware is very large in relation to the size of the image (typically a few cm in each dimension). Typical multiview displays, often based on optical masks or lenticular lens arrays, show multiple 2D images in multiple zones in space. They support multiple simultaneous viewers, but at the cost of restricting them to be within a limited viewing angle. Optical masks cause significant light loss when there are more than two views, and the barrier structure becomes visible as the number of views increases. On the other hand, lenticular displays magnify the pixel matrix of the projecting devices, creating dark zones between viewing slots. The Cambridge multiview display is a classic design in this area [8], and a number of manufacturers (Philips [23], Sharp [25], Opticality [19], Samsung, Stereographics, Zeiss) produce monitors based on variations of this technology. The technology is mature, and, for instance, Spatial View Inc.¹ tried to introduce in the market a medical workstation equipped with an autostereoscopic monitor and an advanced vision based interaction system. A 3D stereo effect is obtained when the left and right eyes see different but matching information. However, the small number of views of multiview systems based on masks or lenticulars generates crosstalks and discontinuities upon viewer motion [7].

The display used in this work [2], produced by Holografika², uses the distributed image generation approach of projector-based multiview technology, but removes some of the intrinsic optical limitations, as it offers a fully continuous blend among views thanks to the light shaping capabilities of a holographically recorded screen.

2.2 High resolution light field display

In the display employed in this work, projectors are densely arranged in a horizontal linear array behind the screen, each one projecting a specific image onto the holographic screen to build up a light field (see figure 1 left). Mirrors located at the sides of the display reflect back onto the screen the beams that would otherwise be lost, creating in this way virtual projectors that increase the display field of view. Each projector emits light beams toward a subset of the points of the holographic screen, so that each screen point is hit by multiple light beams coming from different projectors. The screen, a holographically recorded, randomized surface relief structure, performs selective directional transmission of light beams. Horizontally, the surface is sharply transmissive, in order to maintain a sub-degree separation between views. Vertically, the screen scatters widely, so that the projected image can be viewed from essentially any height. The angular light distribution profile introduced by the holographic screen is characterized by a wide plateau and steep Gaussian slopes precisely overlapping in a narrow region in the horizontal direction. This results in a homogeneous light distribution and continuous 3D view with no visible crosstalk within the field of depth determined by the angular resolution. A full parallax system could be created using a screen with narrow transmission profiles both in the horizontal and vertical direction, as in the improved system presented in [26]. This solution provides in principle a more faithful light field reconstruction, but requires the generation of a much larger number of rays to achieve the same spatial accuracy, which makes it currently practical only for very small image areas or narrow fields of view. Since humans perceive depth using horizontally offset eyes and move their viewpoint more easily from side to side than up and down, the horizontal parallax only approach is considered adequate for most applications and provides a consistent speed-up in computation. As we will see, however, the removal of one degree of freedom in light field reproduction imposes constraints both on the rendering techniques employed and on the design of the interactive applications.

2.3 Projecting graphics on the light field display

The first step to build a rendering pipeline on the display consists in determining where 3D points should be drawn on a given projector to produce a perspective correct image for the viewer. However, the linear perspective is not sufficient, because it ignores the transformation performed by the holographic screen. Since the screen is selective only in the horizontal direction, but scatters widely in the vertical one,

¹ www.healthcare.spatialview.com

² www.holografika.com

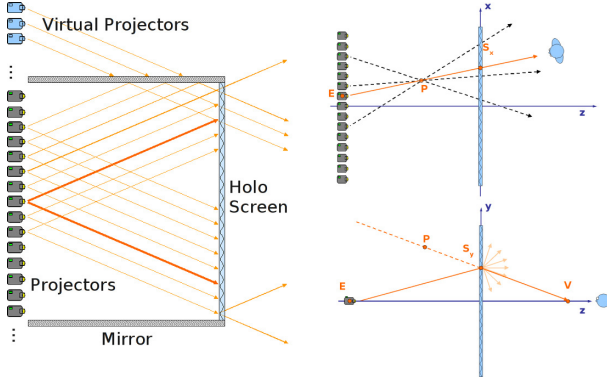


Fig. 1 Display concept. Left: Each projector emits light beams toward a subset of the points of the holographic screen. Top right: horizontally, the screen is sharply transmissive and maintains separation between views. Bottom right: vertically, the screen scatters widely, so that the projected image can be viewed from essentially any height, and the MCOP approximation chooses the ray passing through a virtual observer position.

the displayed light field’s dimensionality is reduced, and the application must decide how to deal with the missing degree of freedom. As explained in [1], we use a multiple center of projection (MCOP) approach (see figure 1 right). In order to provide a full perspective effect, the vertical viewing angle must be known, which amounts to fixing the viewer’s height and distance from screen, while leaving lateral motion completely free. In our coordinate system we place the origin in the screen center, with the y axis in the vertical direction, the x axis pointing to the right, and the z axis pointing out of the screen. Each projector is modeled by a pinhole emitter with origin at $\mathbf{E} = (E_x, E_y, E_z)$, projecting an image on the plane $z = 0$. The projected image geometry is defined by a 2D rectangle $\mathbf{R}^-, \mathbf{R}^+$ orthogonal to the Z axis. Virtual projectors generated by the lateral mirrors are treated as the real ones. The projection of a point \mathbf{P} on the screen \mathbf{S} for a given emitter \mathbf{E} can be computed for the x coordinate by intersecting the ray originating from the emitter \mathbf{E} with the screen plane at $z = 0$ and for the y coordinate by intersecting it with the ray reaching the virtual viewer eye positioned at coordinates $\hat{\mathbf{V}}$:

$$S_x = E_x - E_z \cdot \frac{E_x - P_x}{E_z - P_z} \quad S_y = \hat{V}_y - \hat{V}_z \cdot \frac{\hat{V}_y - P_y}{\hat{V}_z - P_z} \quad (1)$$

The solution is exact for all viewers at the same distance from screen and height as the virtual observer and proves in practice to be a good approximation for all other viewing positions in the display workspace. In this way, given a display size and a sufficiently large viewing angle, it is possible to build an effective multi-user, collaborative tool. Another solution could also consist by tracking the user position in order to properly compute the projected geometries. However, this solution hardly scales with the number of users,

and it may suffer from artifacts due to the synchronization lags involved by the tracking system.

The solution presented above is exact only for an idealized pinhole projection system. The remaining error is corrected by applying a 2D warp that moves pixels to the correct coordinates. For geometric calibration, a classic two step approach is employed, where position and frustum of each projector are found through parametric optimization of an idealized pinhole model and any remaining error is corrected by encoding the correction in the 2D warping coefficients.

3 Volume rendering on the spatial light field display

In this section, we briefly describe our method to drive the light field display in order to get scalable interactive volume ray casting visualization of huge volumetric datasets. Our technique is an enhancement of the single pass out-of-core raycaster presented in [11], adapted to work on light field displays similarly to [1].

The overall approach is based on a GPU raycaster, which follows rays generated by a MCOP projection model, while adaptively sampling prefiltered versions of the dataset at resolutions matching the varying spatial accuracy of the display. Modern diagnostic modalities produce huge datasets that may not be directly loaded into the GPU memory for the rendering pipeline. In order to allow the volume rendering of very large datasets, we rely on an adaptive technique based on the decomposition of a volumetric data set into small cubical bricks, which are then organized into an octree structure maintained out-of-core. The octree contains the original data at the leaves, and a filtered representation of children at the inner nodes. Each node also stores the range of values, as well as high quality precomputed gradients. In order to efficiently support linear interpolation, we replicate one layer of neighboring samples at each brick boundary. The octree is stored in an out-of-core structure, based on Berkeley DB, and data is losslessly compressed with the LZO compression library.

The system has been developed in order to fulfill all requirements involved in the analysis of high quality and high resolution medical and anatomical data. Volume data are represented as 16 bit scalars, and 32 bit gradients. The gradients are precomputed by employing high quality 5x5x5 Sobel filtering. The system is also able to manage and render segmented datasets represented as labelled voxels. In that case, the precomputation of levels of detail is modified to choose for each value the most popular label instead of the average.

At runtime, a working set of bricks is generated and incrementally maintained on CPU and GPU memory by asynchronously fetching data from the out-of-core octree. The

working set is created by an adaptive loader on the basis of the current object position in relation to the screen and transfer function. It is important to note that the 3D display design has consequences not only on the projection equation but also imposes limits on spatial resolution that depend on depth. The size r of the smallest voxel that can be reproduced depends on the distance P_z of its center from the screen, and can be approximated by

$$r(P_z) = r_0 + 2P_z \tan\left(\frac{\alpha}{2}\right) \quad (2)$$

where α is the beam angular size, and r_0 is the pixel size on the screen surface.

In contrast to [11], level of detail selection is observer independent, and our adaptive loader strives to match rendered data resolution with display resolution by stopping octree refinements as soon as the voxel size matches the display resolution. Once the current working set is defined, a compact indexing structure, which spatially organizes the current working set into an octree hierarchy, is encoded in a small texture. This structure spatially organizes the leaves of the current view dependent representation into an octree with neighbor pointers. The inner nodes of this structure simply contain pointers to children, and only the leaves refer to volume data nodes stored in the memory pool. The spatial index structure is exploited by an efficient stackless GPU raycaster, which is able to compute the volume rendering integral by enumerating non-empty bricks in front to back order, adapting sampling density to brick resolution, and stopping as soon as the accumulated opacity exceeds a certain threshold, updating both the framebuffer and the depthbuffer. The prototype volume ray caster implements a number of composition strategies, including Direct Volume Rendering with a Phong illumination model, boundary enhancement and view dependent transparency [5]. When dealing with labeled data, trilinear filtering is substituted with nearest neighbor to preserve label values.

In order to drive the light field display the scene is rendered once per projector view. In our previous work [1], the volume raycaster worked using the idealized pinhole model, and small nonlinear view and color distortions were corrected by streaming the rendering result, stored in a texture, through a fragment shader that warped the geometry and modified colors thanks to per-pixel lookup tables stored as precomputed textures. In this work, we have implemented the full nonlinear projection in the ray setup stage, by encoding color and geometry warping as cubic polynomials.

The resulting system is capable to interactively explore extremely large datasets on light field displays (see figure 2 for some examples).

4 System implementation and interaction design

4.1 Hardware and software system setup

The software system consists of an integrate applications built around a multiresolution volume processing and rendering framework written in C++ and OpenGL, a set of Cg shaders that implement the basic raycasting engine, and a number of shader functions that implement different composing techniques. Data loading and DICOM connectivity is implemented using the OFFIS DCMTK library allowing the access to standard radiology PACS.

The system discussed in this work uses a 26" HoloVizio display by Holografika. The display hardware employed here is capable of visualizing 7.4M beams/frame by composing optical module images generated by 96 fast 320x240 LCD displays fed by FPGA input processing units that decode an input DVI stream. The display is fed by 4 DVI channels working at 1280x1024 at 75Hz. Each 1280x1024 frame collects 16 320x240 projector images, plus a color encoded header in the top rows that encodes the ids of the projectors that have to be updated. A full 3D frame is created by generating all the projector images into the frame buffer. This is obtained by employing a solution in which each DVI channel is fed by a graphic node (PC equipped with a NVIDIA 8800 GTX graphics board), which is charged with filling a subset of the projectors. The on-screen 2D pixel size of the display is $s_0 = 1.25\text{mm}$, and the angular accuracy is 0.8° . The screen width is $L = 500\text{mm}$. and its horizontal viewing angle is $\theta = 50^\circ$.

Even though the display has some limitations in terms of screen size, luminance, and equivalent 2D pixel count, the overall system is already capable of dynamically generating compelling 3D images, and, as we will see in the following sections, the system can already be used as a working testbed for collaborative medical data analysis.

4.2 Visualization/user areas and light field accuracy

As stated before, our system is conceived as a prototype multiuser 3d radiology workstation. In a typical usage scenario, a number of users move in front of the display interacting with the virtual objects, at a distance that they consider comfortable for seeing objects and pointing at them. We fixed this distance to $\hat{V}_z = 1000\text{mm}$ and used it to define the expected user position in the projection model (the other components $\hat{V}_x = 0$ and $\hat{V}_y = 0$ are set in the screen center). These parameters put some limits on the user area: in fact, the minimum distance at which the screen can be observed is $z_{near} = \frac{L}{2} \cot\left(\frac{\theta}{2}\right) = 536\text{mm}$. The horizontal range where

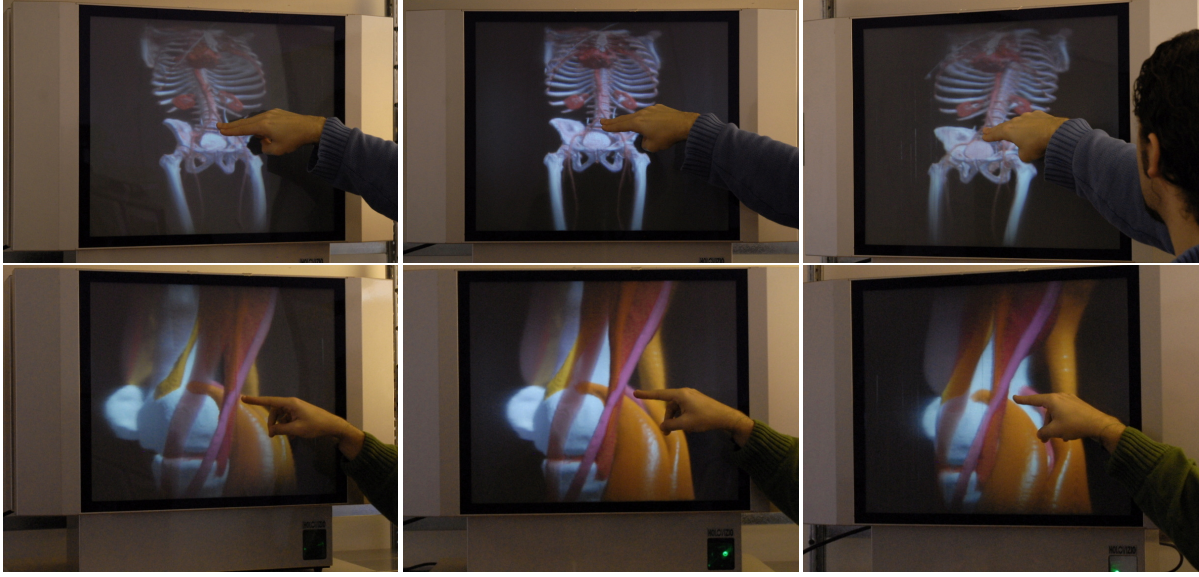


Fig. 2 Inspecting raw and segmented datasets on the light field display. Top: Raw whole body contrast CTA acquired on a 16 detector CT scanner (source: Radiology Department, University of Geneva). The volume has a resolution of $512 \times 512 \times 1559$ with 16 bit/sample. Bottom: segmented leg reconstructed from MRI acquisitions (source: MiraLab, Geneva). The volume has a resolution of $404 \times 474 \times 2050$. Pictures are taken with a hand-held camera at different viewing angles, in order to highlight the horizontal parallax of the light field display. Images are rendered at about 6.5fps.

the screen center is viewable at the expected user distance is $2\hat{V}_z \tan(\frac{\theta}{2}) = 933\text{mm}$, sufficiently large for collaborative analysis.

Another important factor to consider is that voxel size of the virtual object is not constant, but varies with the distance of the object from the screen according to equation 2, with $r_0 = 1.25\text{mm}$ and $\alpha = 0.8^\circ$. It is thus better to limit the depth range in which to manipulate objects. For instance, if we assume that the floating object is put in a bounding box extended 100mm towards the observers, the resolution at the margin of the box is degraded to $r = 4.0\text{mm}$. We have found that putting a clip plane at that depth position is a very natural and effective choice to limit the object extent. It is clear that if the volume resolution is higher than $r(P_z)$, the represented object should be scaled up in order to have a visualization preserving the original detail.

4.3 Projective error

As we stated before, our system is designed to be used as a collaborative radiological workstation. This means that it should allow different users to see the same object floating in the 3D space with a spatially coherent representation. As seen in Sec. 2, the display generates only horizontal parallax, and our MCOP approximation produces exact results for points not lying on the screen plane only if the real viewer's Y and Z coordinates are exactly the same as those of the assumed virtual viewer.



Fig. 3 Projective errors. Left: picture taken from a position at prescribed distance from screen and height from ground ($V_y = 0$, $V_z = 1000\text{mm}$). Center: picture taken from a position much closer to the screen but at standard height ($V_z = 500\text{mm}$). Right: picture taken from a position closer to the screen and nearer to the ground ($V_y = -300\text{mm}$, $V_z = 900\text{mm}$). Note in the last image the slight apparent model rotation due to the lack of vertical parallax.

Figure 3 shows the typical effects of the error introduced when users get closer to the screen, or look at the scene from a lower height. The effect of the projective error is perceived as an angular distortion, which means that the scene appears as horizontally rotated by a certain angle. This effect, as we will see in Sec. 5, does not cause relevant problems in the spatial analysis and in performing complex spatial tasks, because it varies smoothly with the volume coordinates and with user position. However, problems can arise during collaborative work, due to the fact that the same voxel could be seen in slightly different spatial position by different users. For this reason, we found it necessary to quantify this projection error and define a working area where the distortion can be considered negligible (Fig. 4). It should be noted that this limitation appears only if one wants to ensure overlap between physical objects (e.g., user's fingers) and synthe-

sized ones. It does not apply when computer-mediated tools are used (e.g., tracked devices controlling a virtual 3D cursor).

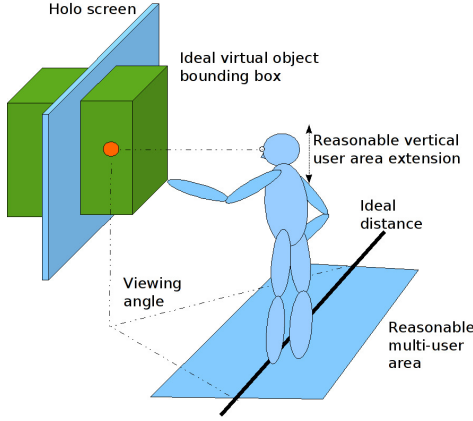


Fig. 4 User/object areas for the optimal use of the system. The necessity of collaborative use and the depth range of the display suggest bounds for virtual object and user positions.

In order to quantify the distortion error, we must use equation (1), which estimates the y screen coordinate based on a virtual viewer position. Distortion artifacts appear when the y or z coordinates of the real viewer are different from the virtual viewer ones. The estimated \hat{S}_y coordinate, computed by fixing the viewer at an expected position \hat{V} , is

$$\hat{S}_y = \hat{V}_y - \hat{V}_z \cdot \frac{\hat{V}_y - P_y}{\hat{V}_z - P_z}, \quad (3)$$

while the real S_y coordinate, computed by considering the real viewer position V , is

$$S_y = V_y - V_z \cdot \frac{V_y - P_y}{V_z - P_z} \quad (4)$$

We want to define a region around the expected user position $\hat{V} = (0, 0, 1000\text{mm})$ where the projection error of volume points (P_x, P_y, P_z) in a bounded object region is small. The object region is limited by a distance from the screen not greater than 100mm and a y distance from the screen center not greater than 150mm. These bounds come from the depth range of the display and the maximum dimension of visible objects given the screen size. We want to loosely constrain the user position in a region where the vertical distance between corresponding points seen by that user and ideal user is less than $\frac{\epsilon}{2}$, where ϵ is the maximum pointing error accepted, that we set to 10mm. Since the maximum error is made at the farthest distance from screen, we obtain a condition on the screen projection error $\Delta S_y = \|\hat{S}_y - S_y\| < 5.55\text{mm}$.

It is then easy to compute, for instance, the displacement along the z axis allowed for users with y coordinate equal

to 0, i.e., of standard height: by putting in equations 3 and 4 the coordinates P_y, P_z of the farthest point of the object region $(150\text{mm}, 100\text{mm})$, $\hat{V} = (0, 0, 1000\text{mm})$, $V_y = 0$, we obtain from the condition $\|\hat{S}_y - S_y\| < 5.55\text{mm}$ that the optimal workspace is the region between $V_z = 775\text{mm}$ and $V_z = 1449\text{mm}$. The condition of unambiguous object pointing determines therefore the user area that is sufficiently large to support collaborative work.

If viewers do not look from $V_y = 0$, i.e., if they are not of standard height, it is also possible to evaluate the distortion with the same method used before. In this case, we obtain a allowed y range varying from $\pm 75\text{mm}$ for $V_z = 775\text{mm}$ to $\pm 155\text{mm}$ for $V_z = 1449\text{mm}$, sufficiently extended to allow people of different heights to use the system together.

4.4 User interaction

User interaction is based on two different interfaces: a standard windows system on a control workstation where a user can select datasets and transfer functions and start the spatial rendering, and a remote control with a 6DOF tracking system included. Users examining the volume representation in the interaction area in front of the screen do not require particular navigation tools, because they can move around the virtual object and examine it. This is a relevant advantage compared with the classical navigation of standard 3D workstations: it is evident that, if useful information can be perceived without need for interaction, understanding will be in general more rapid, since interaction via a computer device adds to cognitive load and always takes more time than making an eye or head movement. Therefore, displacement of the virtual volume from the ideal position is usually not always necessary.

However, a control of position, scale and orientation of the volume representation is still useful for gross manipulation, for example, to zoom and place a part of the volume to be analyzed in the optimal position to obtain visual accuracy. In the current setup, we allow this motion both from the control workstation, using a 2D mouse, and directly in 3D, using 6DOF trackers handled by the OpenTracker reconfigurable tracking subsystem. In the current setup, we mostly use a Logitech 3D mouse. We are also testing low-cost devices like the Nintendo Wii remote game controller (including a 6DOF accelerometer and an IR tracker). The remote control is also useful to allow the user to change visualization options (e.g. windowing, transfer functions) during the examination. In this case, in addition to simple 3D menus handled with the 3D device, we have implemented remote control of the rendering application by a tablet PC or a palmtop computer, using web-based interprocess communication.

5 Perceptual evaluation

The human visual system makes use of a wide variety of cues in understanding a 3D scene. The goal of the experiments described here is to show that the light field display enhances understanding by providing cues not available in traditional display systems.

5.1 Depth cues analysis

First of all, in order to demonstrate that the light field display is able to provide correct and reliable stereo cues, we replicated the tests performed by Kersten et al [14]. We found that the light field system is able to provide appropriate stereoscopic rendering. Eleven subjects obtained success rates in the discrimination of the rotation direction of a cylinder filled with Perlin noise equal to 83% for MIP visualization and 91% for X-RAY visualization over 10 trials. However, ideally the contributions that disparity-based cues make to depth perception should be measurable in the absence of those cues to 3D that are usually available in 2D displays (e.g. object occlusion, surface shading, perspective foreshortening and texture gradients) or other stereo based display systems. Yet it is difficult to find naturalistic scenes in which disparity based cues exist in isolation. We thus decided to resort to an artificial technique employed in the field of visual psychophysics to generate scenes that mask out those visual cues that are not disparity based. Specifically, we designed a synthetic benchmark using random-dot masking in a simplified version of Julesz’s spiral ramp surface: a 3-layer cylindrical wedding cake model that subjects viewed along its concentric axis (see figure 5 right). The random dot textures block the usual pathways along which vision proceeds: the locations of surface boundaries within the scene are lost in the “cloud of dots”. Vision can then only make sense of the scene after achieving binocular fusion, which gradually reveals the correspondence between the random dots as seen from two different viewpoints, e.g. the left and right eyes. The expectation here is that on the light field display such a scene will be seen almost instantaneously, thus demonstrating the added value that it brings. By adjusting the model’s parameters and converting it to a rectilinear volume, two sets of model stimuli were rendered: one with a uniform large field of depth ($\pm 10\text{cm}$ centered on the display screen) and one where the field was almost flat ($\pm 1\text{cm}$). Eleven, pre-screened, subjects completed four experiments, each consisting of eight trials in a two-interval forced-choice (2IFC) design whereby they indicated in which interval they perceived the greatest field of depth. The experiments tested one eye static, one eye

head-swaying, two eye static, and two eye head-swaying observation in that order. Scores improved also in that order: from 49%, i.e., indistinguishable from a random answer in the binary test, with a monocular static view, to 82% correct scores for the monocular head-swaying test, up to 100% when all cues are available. The results indicate that in the absence of cues normally available in 2D displays the light field rendering system elicits useful stereoscopic and motion parallax depth cues.

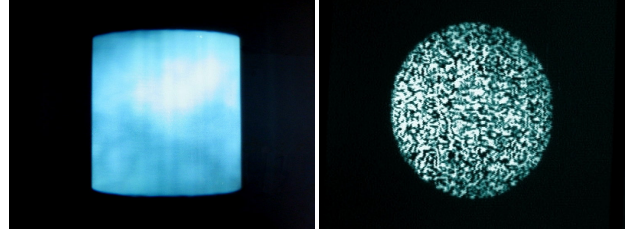


Fig. 5 Depth cues evaluation: left: cylinder filled with Perlin noise; right: 3-layer cylindrical wedding-cake model with random-dot texture.

5.2 Layout discrimination performance

While helpful for describing display system characteristics and assessing image quality, standard perceptual evaluation metrics provide a limited understanding of the task relevant issues for display design and do not address task specific requirements needed by users. In order to evaluate performance of our spatial light field display, we considered task specific requirements needed by neurosurgeons or physicians involved in the analysis of medical images such as angiography datasets (see figure 6 left). In order to quantify the performance of the system with reference to network structures understanding, we employed a performance test where users were asked to trace complex paths [24]. Specifically, subjects were presented with a scene rendered with a Maximum Intensity Projection technique in which the dataset contains a graph having one of the nodes highlighted in red (see figure 6 right) and asked to determine how many nodes compose the highlighted path. The task was generally found very tricky, since it forces subjects to employ all cues involved in layout discrimination, such as stereopsis, accommodation, convergence, and motion parallax. The test was performed by 14 subjects who never had experience with the light field display. They were confronted with 10 trials, with a graph randomly generated and containing 20 paths, each one containing 2 to 6 nodes. Three cases were considered: static scene in a light field display, static scene in a 2D display, and free mouse interaction in a 2D display. Users had no timing constraints.

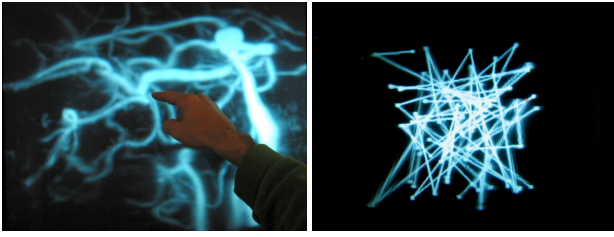


Fig. 6 Path tracing performance tests. Left: angiography data inspection with the light field display. Right: synthetic test scene employed for evaluating path tracing performance.

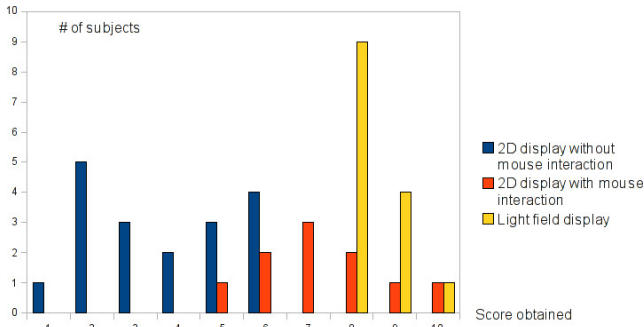


Fig. 7 Path tracing performance histograms. The x axis represents the score over 10 trials, while the y axis represents the number of subjects who achieved that score. Blue (dark) bars: results obtained with light field display. Red (intermediate) bars: results obtained with a 2D desktop display and free mouse interaction enabled. Yellow (light) bars: results obtained with a 2D desktop display and mouse interaction disabled.

The histograms in figure 7 show the performance results, for the three conditions considered. For each diagram, the x axis represents the number of correct answers, while the y axis represents the number of subjects who obtained that score. The results highlight that the performance is clearly higher for spatial light field display (error rate of about 15%) as opposed to the 2D display with mouse interaction enabled (error rate of about 33%). Finally, the 2D desktop without possibility of interaction was confirmed to be completely useless (error rate bigger than 70%).

5.3 Preliminary evaluation with end users

In order to informally test the diagnostic capabilities of the system, collaborative discussion sessions were also performed on the light field display. Three expert radiologists and physicians discussed the volume visualization of some anonymous datasets. Datasets were rendered by employing the Maximum Intensity Projection or X-ray volume rendering technique (figure 8). These techniques are depth-oblivious, and do not provide useful depth cues in static bi-dimensional rendering.



Fig. 8 Collaborative diagnostic sessions Pictures taken from collaborative diagnostic discussion sessions on MIP rendered angiography datasets and X-ray rendered orthopedic datasets.

Three view positions were considered for the discussion: antero-posterior, left posterior oblique, and sagittal view. In all cases, physicians judged that anatomical and pathological information was very easy to interpret from spatial 3D observation on the light field display, with no manipulation. The two posterior cerebral arteries were immediately recognized, and physicians were pleased and surprised by this fact, since the understanding of principal structure in a CTA is a task involving some efforts even for trained radiologists. They reported that this kind of visualization can greatly improve the communication and the operative planning for neurosurgery. Physicians were particularly impressed by the depth perception of vascular structures and by discrimination easiness. The discussion and collaboration was driven largely by the clarity of images, and physicians were able to describe and highlight all vascular structures inside the datasets. The general opinion is that the strength of such a system is related to the collaboration, discussion, and evaluation of clinical cases, since all users can have immediate 3D understanding of anatomy.

6 Discussion

We have described the implementation of a new medical visualization system having several features that are expected to be found in the 3D radiology workstation of the future: a light field display able to provide stereo and motion parallax cues in a setting supporting collaborative use, and advanced volumetric rendering implementation that can create, at interactive speed, high quality light fields from gigavoxel datasets.

Our prototype has to be considered a testbed in a development process aimed at creating a really good 3D light field based radiology workstation in the future. Despite the limitations of current prototypical hardware settings, all the perceptual evaluation tests that we have carried out tend to prove that the system provides correct depth cues, helps in layout discrimination and is clearly superior to two-dimensional displays for path tracing

tasks, common in operations such as the understanding of vascular structures. The first feedback received from physicians and radiologists seems to confirm this fact.

Besides continuing to improve our implementation, we plan to further extend the evaluation of the 3D display capabilities in comparison with other stereo display solutions, with a special focus on the immediacy and persistence of the stereo effect, and to perform an evaluation of the system for diagnostic tasks in a clinical context.

Acknowledgments

This work is partially supported by the Italian Ministry of Research under the CYBERSAR project and by the EU Marie Curie Program under the 3DANATOMICALHUMAN project (MRTN-CT-2006-035763).

References

- Agus, M., Gobbetti, E., Guitián, J.A.I., Marton, F., Pintore, G.: GPU accelerated direct volume rendering on an interactive light field display. *Computer Graphics Forum* **27**(2), 231–240 (2008)
- Balogh, T., Forgacs, T., Agoes, T., Balet, O., Bouvier, E., Bettio, F., Gobbetti, E., Zanetti, G.: A scalable hardware and software system for the holographic display of interactive graphics applications. In: *Eurographics Short Papers Proceedings*, pp. 109–112 (2005)
- Bettio, F., Gobbetti, E., Marton, F., Pintore, G.: Scalable rendering of massive triangle meshes on light field displays. *Computers & Graphics* **32**(1), 55–64 (2008)
- Boucheny, C., Bonneau, G.P., Droulez, J., Thibault, G., Ploix, S.: A perceptive evaluation of volume rendering techniques. In: *Proc. ACM APGV*, pp. 83–90 (2007)
- Bruckner, S., Gröller, M.E.: Style transfer functions for illustrative volume rendering. *Computer Graphics Forum* **26**(3), 715–724 (2007)
- Cossairt, O., Napoli, J., Hill, S., Dorval, R., Favallora, G.: Occlusion-capable multiview volumetric three-dimensional display. *Applied Optics* **46**(8), 1244–1250 (2007)
- Dodgson, N.A.: Analysis of the viewing zone of the Cambridge autostereoscopic display. *Applied Optics: Optical Technology & Biomedical Optics* **35**(10), 1705–1710 (1996)
- Dodgson, N.A., Moore, J.R., Lang, S.R., Martin, G., Canepa, P.: Time-sequential multi-projector autostereoscopic 3D display. *J. Soc. for Information Display* **8**(2), 169–176 (2000)
- Favallora, G., Dorval, R., Hall, D., Napoli, J.: Volumetric three-dimensional display system with rasterization hardware. In: *Proc. SPIE*, vol. 4297, pp. 227–235 (2001)
- Favallora, G.E.: Volumetric 3d displays and application infrastructure. *Computer* **38**(8), 37–44 (2005)
- Gobbetti, E., Marton, F., Iglesias Guitián, J.: A single-pass GPU ray casting framework for interactive out-of-core rendering of massive volumetric datasets. *The Visual Computer* **24**(7–9), 797–806 (2008)
- Huebschman, M., Munjuluri, B., Garner, H.: Dynamic holographic 3-d image projection. *Optics Express* **11**, 437–445 (2003)
- Jones, A., McDowall, I., Yamada, H., Bolas, M.T., Debevec, P.E.: Rendering for an interactive 360 degree light field display. *ACM Trans. Graph* **26**(3), 40 (2007)
- Kersten, M., Stewart, J., Troje, N., Ellis, R.: Enhancing depth perception in translucent volumes. *IEEE Transactions on Visualization and Computer Graphics Journal* **12**(6), 1117–1123 (2006)
- McKay, S., Mair, G., Mason, S., Revie, K.: Membrane-mirror based autostereoscopic display for teleoperation and telepresence applications. In: *Proc. SPIE*, vol. 3957, pp. 198–207 (2000)
- Mora, B., Ebert, D.S.: Instant volumetric understanding with order-independent volume rendering. *Computer Graphics Forum* **23**(3), 489–497 (2004)
- Napoli, J., Stutsman, S., Chu, J.C.H., Gong, X., Rivard, M.J., Cardarelli, G., Ryan, T.P., Favallora, G.E.: Radiation therapy planning using a volumetric 3-D display: PerspectaRAD. p. 680312. *SPIE* (2008)
- Raap, G.B., Koning, A.H., Scohy, T.V., ten Harkel, A.D.J., Meijboom, F.J., Kappetein, A.P., van der Spek, P.J., Bogers, A.J.: Virtual reality 3D echocardiography in the assessment of tricuspid valve function after surgical closure of ventricular septal defect. *Cardiovascular Ultrasound* **5**(8) (2007)
- Relke, I., Riemann, B.: Three-dimensional multiview large projection system. In: *Proc. SPIE*, vol. 5664 (2005)
- Roberts, J.W., Slattery, O.: Display characteristics and the impact on usability for stereo. In: *Proc. SPIE*, vol. 3957, p. 128 (2000)
- St.-Hillaire, P., Lucente, M., Sutter, J., Pappu, R., Sparrell, C.G., Benton, S.: Scaling up the MIT holographic video system. In: *Proc. 5th SPIE Symposium on Display Holography*, pp. 374–380 (1995)
- Stanley, M., Conway, P., Coomber, S., Jones, J., Scattergood, D., Slinger, C., Bannister, B., Brown, C., Crossland, W., Travis, A.: A novel electro-optic modulator system for the production of dynamic images from giga-pixel computer generated holograms. In: *Proc. SPIE*, vol. 3956, pp. 13–22 (2000)
- van Berkel, C., Parker, D., Franklin, A.: Multiview 3d-lcd. In: *Proc. SPIE*, vol. 2653, p. 32 (1996)
- Ware, C., Franck, G.: Evaluating stereo and motion cues for visualizing information nets in three dimensions. *ACM Trans. Graph.* **15**(2), 121–140 (1996)
- Woodgate, G.J., Harrold, J., Jacobs, A.M.S., Moseley, R.R., Ezra, D.: Flat-panel autostereoscopic displays: characterisation and enhancement. In: *Proc. SPIE*, vol. 3957, p. 153 (2000)
- Yang, R., Huang, X., Li, S., Jaynes, C.: Toward the light field display: Autostereoscopic rendering via a cluster of projectors. *IEEE Transactions on Visualization and Computer Graphics* **14**(1), 84–96 (2008)



Marco Agus is a researcher in the Visual Computing group at the Center for Advanced Studies, Research, and Development in Sardinia (CRS4). He holds a Laurea (M.Sc.) degree in Electronics Engineering (1999) and a Ph.D. degree in Mechanical Engineering (2004) from the University of Cagliari, Italy. His research interests include large model rendering of novel displays, volume visualization, haptic and surgical simulations.



José Antonio Iglesias Guitián is a Marie Curie Early Stage Researcher in the Visual Computing group of the Center for Advanced Studies, Research and Development in Sardinia (CRS4). He obtained his Masters degree in Computer Science in 2006 at the University of A Coruna, Spain.



Fabio Bettio is a senior software engineer in the Visual Computing group at the Center for Advanced Studies, Research, and Development in Sardinia (CRS4). Research, and Development in Sardinia (CRS4). He holds a Laurea (M. Sc.) degree (1994) in Electronics Engineering from the University of Genoa, Italy.



Fabio Marton is a researcher in the Visual Computing group at the Center for Advanced Studies, Research, and Development in Sardinia (CRS4). He holds a Laurea (M.Sc.) degree (1999) in Computer Engineering from the University of Padua, Italy. His current research interests include out-of-core data processing, multiresolution modeling and time-critical rendering.



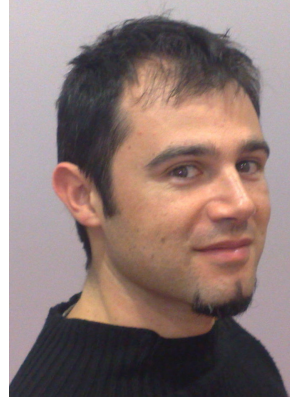
Andrea Giachetti is Associate Professor of Computer Science at the University of Verona, Italy, and holds a research position with the Center for Advanced Studies, Research, and Development in Sardinia (CRS4). He received a M.Sc. degree (1993) and a Ph.D. degree in Physics (1997) from the University of Genoa, Italy. His research interests cover computer vision, medical image analysis, image segmentation and 3D reconstruction.



Jonas Nilsson carried out this work during an internship in the Visual Computing group at the Center for Advanced Studies, Research, and Development in Sardinia (CRS4). The internship was part of his Master of Science degree program at Linköping University, Sweden.



Enrico Gobbetti is the director of the Advanced Computing and Communications Program and of the Visual Computing group at the Center for Advanced Studies, Research, and Development in Sardinia (CRS4). His research spans many areas of computer graphics and is widely published in major journals and conferences. Enrico holds an Engineering degree (1989) and a Ph.D. degree (1993) in Computer Science from the Swiss Federal Institute of Technology in Lausanne (EPFL).



Giovanni Pintore is a software engineer in the Visual Computing group at the Center for Advanced Studies, Research, and Development in Sardinia (CRS4). He holds a Laurea (M. Sc.) degree (2002) in Electronics Engineering from the University of Cagliari, Italy.